IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of current version December 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3041166

EDITORIAL

IEEE ACCESS SPECIAL SECTION EDITORIAL: SCALABLE DEEP LEARNING FOR BIG DATA

Deep learning (DL) has emerged as a key application exploiting the increasing computational power in systems such as GPUs, multicore processors, Systems-on-Chip (SoC), and distributed clusters. It has also attracted much attention in discovering correlation patterns in data in an unsupervised manner and has been applied in various domains including speech recognition, image classification, natural language processing, and computer vision. Unlike traditional machine learning (ML) approaches, DL also enables dynamic discovery of features from data. In addition, now, a number of commercial vendors also offer accelerators for deep learning systems (such as Nvidia, Intel, and Huawei).

Typical deep neural networks (DNNs) require large amounts of data to learn parameters (often reaching millions), which is a computationally intensive process requiring significant time to train a model. As the data size increases and as deep learning models become more complex, it requires more computing power and memory to train an accurate model in a timely manner. Despite existing efforts in training and inference of deep learning models to increase concurrency, many existing training algorithms for deep learning are notoriously difficult to scale and parallelize due to inherent algorithmic interdependencies and the training data. Therefore, there is a need to develop efficient and scalable deep learning frameworks suitable for big data processing and analysis.

This Special Section of IEEE Access on scalable deep learning for big data brings together research contributions from academia and industry which address key challenges in big data processing and analysis using scalable deep learning.

The Call for Papers attracted a lot of attention from the scientific community and received 92 submissions, out of which 28 articles were accepted for inclusion in the Special Section after a thorough review process. Each submission was reviewed by at least two independent referees. The accepted articles cover a range of topics relevant to this Special Section including algorithmic developments to support applications in different domains (such as healthcare, transportation, agriculture, power system, flood monitoring, mine gas monitoring, social networks, cyber security, software engineering, etc.) and using various types of big data (text, images, videos, time-series data, etc.).

In the article "Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks," by Cho *et al.*, the authors describe a simple yet robust algorithm for DNN hyperparameter optimization—DEEP-BO (diversified, early termination-enabled, and parallel Bayesian optimization). When evaluated over six DNN benchmarks, DEEP-BO mostly outperformed well-known solutions including GP-Hedge, BOHB, and the speed-up variants that use median stopping rule or learning curve extrapolation.

In the article "GAN-knowledge distillation for one-stage object detection," by Wang *et al.*, the authors propose a clean and effective knowledge distillation method called generative adversarial networks–knowledge distillation (GAN-KD) for one-stage object detection. The feature maps generated by teacher and student network are employed as true and fake samples, respectively, and generate adversarial training to improve the performance of the student network in one-stage object detection.

In the article "Attention-based dense decoding network for monocular depth estimation," by Wang *et al.*, the authors describe a novel encoder–decoder attention-dense decoding network to solve the local depth detail loss caused by convolution stacking. It takes advantage of the channel–spatial attention module, which captures the dependence between different channels and spatial locations by self-attentions. It also introduces a dense decoding module to capture more massive and denser attention features. A distance-aware loss function that pays more attention to long-distance objects (i.e., objects that are further away in an image) is also introduced.

In the article "R3MR: Region growing based 3D mesh reconstruction for big data platform," by Li *et al.*, a novel region growing based 3-D mesh reconstruction method for a big data platform is introduced. This approach reconstructs the classified data points from simple to complex by the rational principle of optimal selection and the use of the inner edge adjacency list. Experimental results show that the proposed method can accurately reconstruct the surface shape of the point cloud model and could reflect the detailed features of the model more naturally.

The article by Yasir *et al.*, "TRICE: Mining frequent itemsets by iterative TRimmed transaction LattICE in sparse big data," introduces a novel method to mine frequent itemsets by iterative TRimmed transaction lattICE (TRICE) for processing and analyzing sparse data sets. The proposed method shows better performance than the existing

algorithms including HARPP, FP-Growth, optimized SaM, and optimized RElim.

In the article ''D-GENE: Deferring the GENEration of power sets for discovering frequent itemsets in sparse big data,'' Yasir *et al.* present D-GENE, a technique that optimizes TRICE by introducing a deferred iterative trimmed transaction lattice (ITTL) generation mechanism. D-GENE suspends the process of ITTL generation until the completion of a transaction pruning phase. The deferral strategy enables D-GENE to generate ITTLs of similar trimmed transactions once.

In the article ''A cloud-based framework for machine learning workloads and applications,'' by López García *et al.*, a distributed architecture called DEEP-Hybrid-DataCloud is proposed to provide machine learning practitioners with a set of tools and cloud services that cover the whole machine learning development cycle. This approach includes support for models creation, training, validation, and testing to the models being used as a service, as well as their sharing and publication. It allows transparent access to existing e-infrastructures, effectively exploiting distributed resources for the most compute-intensive tasks. It also provides scientists with a set of cloud-oriented services to make these models publicly available by adopting a serverless architecture and a DevOps approach, and allowing an easy share, publish, and deploy of the developed models.

In the article ''Interpretability analysis of heartbeat classification based on heartbeat activity's global sequence features and BiLSTM-attention neural network,'' by Li *et al.*, a new framework is introduced for ECG heartbeat classification based on BiLSTM-attention neural network model with heartbeat activity's global sequence features for accurate heartbeat classification. This framework can simulate the thinking process of medical experts in diagnosing diseases, and it automatically learns the characteristics of heartbeat categories. The significance of this study is to provide better clinical monitoring, diagnosis, and treatment for heart disease patients.

The article by Guan *et al.*, ''A non-contact paraparesis detection technique based on 1D-CNN,'' presents a noncontact wireless sensing method based on RF signals to detect paraparesis. A 1D-CNN model is designed to automatically extract features of wireless signals and detect paraparesis. The proposed system aims to reduce the burden on doctors and improve work efficiency.

The article ''Uncertainty assisted robust tuberculosis identification with Bayesian convolutional neural networks,'' by Ul Abideen *et al.*, presents a solution for TB identification using a Bayesian-based convolutional neural network (B-CNN). It deals with the uncertain cases that have low discernibility among the TB and non-TB manifested CXRs.

The article ''A deep learning model based on concatenation approach for the diagnosis of brain tumor,'' by Noreen *et al.*, proposes a deep learning method of multilevel feature extraction and concatenation for early diagnosis of brain tumors.

The article ''Epileptic seizures prediction using deep learning techniques,'' by Usman *et al.*, describes a seizure prediction system that employs deep learning methods. The system includes preprocessing of scalp EEG signals and automated feature extraction using convolution neural networks and classification with the support of vector machines.

The article ''Deep convolution neural network for big data medical image classification,'' by Ashraf *et al.*, proposes a novel deep convolution network-based approach that can assist doctors and physicians in making reasonable decisions.

The article by Guo *et al.*, ''A deep learning based fault diagnosis method with hyperparameter optimization by using parallel computing,'' proposes an intelligent fault diagnosis method of rolling bearings based on deep belief network (DBN) with hyperparameter optimization using parallel computing.

The article ''Deep forest regression for short-term load forecasting of power systems,'' by Yin *et al.*, proposes deep forest regression for the short-term load forecasting of power systems. Deep forest regression includes two procedures, i.e., multigrained scanning procedure and cascade forest procedure.

The article by Kim *et al.*, ''AI-IDS: Application of deep learning to real-time web intrusion detection,'' describes an optimal CNN-LSTM model based on SFL and successfully applied payload-level deep learning techniques in a high-performance computing environment. The AI-IDS distinguishes between normal and abnormal traffic on HTTP traffic that could not be detected in legacy signature-based NIDS because AI-IDS can formalize unknown patterns and help write or improve signature-based rules for new vulnerabilities, variants, and bypass attacks.

The article ''SPSR-FSPG: A fast simulative password set generation algorithm,'' by Zhang *et al.*, introduces a fast simulative password set generation model by deeply mining the structural characteristics of passwords in the data set, denoted as SPSR-FSPG. The algorithm uses probability context-free grammar to model the structure of the password, and constructs a string generation model based on a recurrent neural network to generate different types of strings, so as to learn the character composition of the password in the original data set.

The article ''Scalable and secure big data IoT system based on multifactor authentication and lightweight cryptography,'' by Atiewi *et al.*, introduces a secure cloud–IoT environment using multifactor authentication and lightweight cryptography schemes. The proposed method splits IoT devices into sensitive and nonsensitive devices. Sensitive device data are divided into two and encrypted using the RC6 and Fiestel encryption algorithms. These data are stored in a private cloud to provide high security via a gateway device. By contrast, nonsensitive device data are encrypted using AES and stored in a public cloud via a gateway device. Multifactor authentication is provided by the Trusted Authority.

The article ''Scalable mutation testing using predictive analysis of deep learning model,'' by Naeem *et al.*, introduces

a new approach for software mutation testing which extracts features from mutant programs based on mutant killing conditions, i.e., reachability, necessity, and sufficiency, along with mutant significance and test suite metrics to extract features from mutant programs. A deep learning Keras model is designed to predict killed and alive mutants from each program.

The article ''Large-scale text classification using scope-based convolutional neural network: A deep learning approach,'' by Wang *et al.*, proposes a novel large-scale scope-based convolutional neural network (LSS-CNN) for extraction of the most valuable local information of text documents, which is based on scope convolution, aggregation optimization, and max pooling operation. The experimental results show that LSS-CNN can achieve both effectiveness and good scalability on big text data.

The article ''A novel co-training based approach for the classification of mental illnesses using social media posts,'' by Tariq *et al.*, presents a method to classify patients associated with chronic mental illness, i.e., Anxiety, Depression, Bipolar, and ADHD (Attention Deficit Hyperactivity Disorder), based on a cotraining (a type of semisupervised learning approach) technique and the data extracted from Reddit, a well-known network community platform.

The article by Ge *et al.*, ''User topic preferences based influence maximization in overlapped networks,'' presents a method for maximizing influence in overlapping networks based on user interests. An Influence Propagation Model of the Overlay Network (UI-IPM) model is proposed considering both the information transmission characteristics of overlap-ping network nodes and the user's interest characteristics. Based on the UI-IPM model, a heuristic algorithm combined with the greedy algorithm is designed to maximize the impact of overlapping networks (UI-IPM) and achieve Influence Maximization of the Overlay Network (IMON) mining seed nodes.

The article ''A light CNN for end-to-end car license plates detection and recognition,'' by Wang *et al.*, proposes a multi-task convolutional neural network for license plate detection and recognition (MTLPR) with better accuracy and lower computational cost, and also introduces a comprehensive data set of Chinese license plates.

The article ''Multi-feature view-based shallow convolutional neural network for road segmentation,'' by Junaid *et al.*, presents a shallow and robust road segmentation model based on a Multi-feature View-based Shallow Convolutional Neural Network (MVS-CNN).

The article by Khalaf *et al.*, ''IoT-enabled flood severity prediction via ensemble machine learning models,'' describes a new approach for the prediction of water level in association with flood severity using the ensemble model. The proposed approach leverages the latest developments in the Internet of Things (IoT) and machine learning for the automated analysis of flood data that might be useful to prevent natural disasters.

The article ''Research on a mine gas concentration forecasting model based on a GRU network,'' by Jia *et al.*, presents a mine gas concentration prediction model based on gated recurrent units (GRUs) with high accuracy.

The article ''Integration of Google Play content and frost prediction using CNN: Scalable IoT framework for big data,'' by Latif *et al.*, introduces a CNN model approach for frost event prediction based on sensor data collected from the field.

The article ''ECMCRR-MPDNL for cellular network traffic prediction with big data,'' by Dommaraju *et al.*, presents a method based on Expected Conditional Maximization Clustering and Ruzicka Regression-based Multilayer Perceptron Deep Neural Learning (ECMCRR-MPDNL) to achieve higher network traffic prediction accuracy and minimal time with big cellular data.

In conclusion, we would like to thank all the authors who submitted their research articles to our Special Section. We highly appreciate the contributions of the reviewers for their constructive comments and suggestions, and for returning their reviews on time. We also would like to acknowledge the guidance from the Editor-in-Chief and other members of IEEE ACCESS editorial office.

LIANGXIU HAN, *Lead Editor*
Department of Computing and Mathematics
Manchester Metropolitan University
Manchester M1 5GD, U.K.

DAOQIANG ZHANG, *Guest Editor*
College of Computer Science and Technology
Nanjing University of Aeronautics and Astronautics
Nanjing 210016, China

OMER RANA, *Guest Editor*
School of Computer Science and Informatics
Cardiff University
Cardiff CF24 3AA, U.K.

YI PAN, *Guest Editor*
Department of Computer Science
Georgia State University
Atlanta, GA 30302, USA

SOHAIL JABBAR, *Guest Editor*
Department of Computational Science
The University of Faisalabad
Faisalabad 38000, Pakistan

MAZIN YOUSIF, *Guest Editor*
DXC Technology
Tysons, VA 22102, USA

MOAYAD ALOQAILY, *Guest Editor*
Faculty of Engineering
Al Ain University
Al Ain, United Arab Emirates

LIANGXIU HAN received the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2002. She is currently a Full Professor of computer science with the Department of Computing and Mathematics, Manchester Metropolitan University. Having worked in both industry and academia, she has over 18 years of research and practical experience in developing intelligent ICT-enabled software solutions in different application domains, such as health, smart cities, bioscience, cybersecurity, and energy. As a Principal Investigator (PI) or a Co-PI, she has been conducting research in relation to large-scale data processing, big data analysis and mining, and parallel and distributed computing/cloud computing (funded by EPSRC, BBSRC, Innovate UK, Horizon 2020, Royal Society, British Council, Industry, Charity, and so on). She has published over 100 articles in peer-reviewed international journals and book chapters. Her research areas mainly include the development of novel big data analytics and intelligent architectures that facilitate big data analytics, such as parallel and distributed computing and cloud/service-oriented computing/data-intensive computing, as well as applications in different domains using various large data sets, such as biomedical images, environmental sensor, network traffic data, and web documents. She is also an Associate Editor of IEEE ACCESS, a Guest Editor for a number of reputable international journals, including *PPNA, Computational Science*, and *Journal of Medical Systems*, and the chair (or the co-chair) of a number of international conferences/workshops in the field. She has been invited to give a number of keynotes and talks on different occasions, including international conferences.

DAOQIANG ZHANG received the B.Sc. and Ph.D. degrees in computer science from the Nanjing University of Aeronautics and Astronautics, China, in 1999 and 2004, respectively. In 2004, he joined the Department of Computer Science and Engineering, NUAA, as a Lecturer, where he is currently a Professor. His research interests include machine learning, pattern recognition, data mining, and medical image analysis. In these areas, he has published over 200 scientific articles in refereed international journals, such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON MEDICAL IMAGING, IEEE TRANSACTIONS ON IMAGE PROCESSING, *NeuroImage*, *Human Brain Mapping*, and *Medical Image Analysis* and in conference proceedings, such as IJCAI, AAAI, NIPS, CVPR, MICCAI, KDD, with over 12 000 citations by Google Scholar. He was nominated for the National Excellent Doctoral Dissertation Award of China in 2006. He received the best paper award or the best student award from several international conferences, such as PRICAI 2006, STMI 2012, and BICS 2016. He has served as a Program Committee Member for several international conferences, such as IJCAI, AAAI, NIPS, MICCAI, SDM, PRICAI, and ACML. He is a member of the Machine Learning Society of the Chinese Association of Artificial Intelligence (CAAI) and the Artificial Intelligence and Pattern Recognition Society of the China Computer Federation (CCF).

OMER RANA received the Ph.D. degree in neural computing and parallel architectures from Imperial College London in 1998. He is currently a Professor of performance engineering. His research interests include high-performance distributed systems and data analytics. He has published over 310 papers in peer-reviewed international conferences and journals. He serves on the Editorial Board of IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, *ACM Transactions on Internet Technology*, and *ACM Transactions on Autonomous and Adaptive Systems*. He has served as a Co-Editor for a number of journals, including *Concurrency: Practice and Experience* (John Wiley), IEEE TRANSACTIONS ON SYSTEM, MAN, AND CYBERNETICS: SYSTEMS, and IEEE TRANSACTIONS ON CLOUD COMPUTING.

YI PAN received the B.Eng. and M.Eng. degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and the Ph.D. degree in computer science from the University of Pittsburgh, USA, in 1991. He has been a Regents' Professor and the Chair of the Computer Science Department, Georgia State University, since January 2006. He was an Interim Associate Dean and the Chair of the Biology Department, from 2013 to 2017. In 2000, he joined Georgia State University, where he was promoted to Full Professor in 2004, named a Distinguished University Professor in 2013, and designated as a Regents' Professor (the highest recognition given to a faculty member by the University System of Georgia) in 2015. His profile has been featured as a distinguished alumnus in both the Tsinghua Alumni Newsletter and the University of Pittsburgh CS Alumni Newsletter. He has published more than 450 articles, including over 250 journal articles with more than 100 articles published in IEEE/ACM Transactions/journals. In addition, he has edited or authored 43 books. His work has been cited more than 13 500 times based on Google Scholar. His current H-index is 70. His current research interests mainly include bioinformatics and health informatics using big data analytics, cloud computing, and machine learning technologies. He was a recipient of many awards, including one IEEE Transactions Best Paper Award, five IEEE and other international conference or journal best paper awards, four IBM faculty awards, two JSPS senior invitation fellowships, the IEEE BIBE Outstanding Achievement Award, the IEEE Outstanding Leadership Award, the NSF Research Opportunity Award, and the AFOSR Summer Faculty Research Fellowship. He has organized numerous international conferences and delivered keynote speeches at over 60 international conferences around the world. He has served as the Editor-in-Chief or an Editorial Board Member for 20 journals, including seven IEEE TRANSACTIONS. He is also serving as an Associate Editor-in-Chief for IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS.

SOHAIL JABBAR served in different academic and managerial positions at National Textile University, COMSATS University Islamabad, and Bahria University, Islamabad, where he also headed different research groups. He is currently a Postdoctoral Fellow with the CfACS IoT Lab, Manchester Metropolitan University. He is also the Head of the Network Communication and Media Analytics Research Group, National Textile University. He has authored two book chapters and has published over 100 research articles in prestigious journals. He has been engaged in many national- and international-level projects. He is on collaborative research with renowned research centers and institutes around the globe on various issues in the domains of the IoT, WSN, and blockchain. He is a Guest Editor of special issues in leading journals of his domain. He is also engaged as a TPC member/chair in many conferences. He is also a Guest Editor of IEEE ACCESS, *Concurrency and Computation Practice and Experience* (Wiley), *Future Generation Computer Systems* (Elsevier), *Peer-to-Peer Networking and Applications* (Springer), *Journal of Information and Processing System* (KIPS), *Cyber-Physical System* (Taylor & Francis), IEEE WIRELESS COMMUNICATIONS (IEEE Communication Society), and *IEEE Internet of Things Magazine*.

MAZIN YOUSIF received the master's and Ph.D. degrees from Pennsylvania State University in 1987 and 1992, respectively. He was the Vice President of Digital Transformation and the Chief Technology Officer for the Royal Dutch Shell Global account at T-Systems International. He also held executive positions at IBM Corporation and Intel Corporation. He is currently the Chief Technology Officer of energy, utility, mining, oil, and gas with DXC Technology. He was an Adjunct Professor for several universities, including Duke University, NCSU, OGI, and Arizona University. He has served as the general chair or the program chair for many conferences and serves on the editorial board of many journals. He is a frequent speaker in academic and industry conferences on topics related to cloud computing, the IoT, big data, and autonomic computing. He has published extensively. He was an IEEE Distinguished Visitors Program Speaker. From 2008 to 2016, he chaired the Advisory Board of the European Research Consortium for Informatics and Mathematics (ERCIM). From 2014 to 2018, he was the Editor-In-Chief of the *IEEE Cloud Computing Magazine*. He founded the NSF Industry/University Cooperative Research Center for Autonomic Computing.

MOAYAD ALOQAILY (Member, IEEE) received the M.Sc. degree in electrical and computer engineering from Concordia University, Montreal, QC, Canada, in 2012, and the Ph.D. degree in electrical and computer engineering from the University of Ottawa, Ottawa, ON, Canada, in 2016. He was an Instructor with the Systems and Computer Engineering Department, Carleton University, Ottawa, in 2017. He has been working with Gnowit Inc. as a Senior Researcher and a Data Scientist, since 2016. He has been the Managing Director of xAnalytics Inc., Ottawa, since 2019. He is currently with the Faculty of Engineering, Al Ain University, United Arab Emirates. He is also managing a SIG on Blockchain Application with IEEE TC CommSoft. His current research interests include the applications of AI and ML, connected and autonomous vehicles, blockchain solutions, and sustainable energy and data management. He is an ACM Member. He has chaired and co-chaired many IEEE conferences and workshops, including BCCA2020, AdHocNets2020, PEDISWESA-ISCC2020, ITCVT-NOMS2020, E2NIoT-IWCMC2020, ICCN-INFOCOM19, AICSSA19, and BAT-FMEC19-20. He has served as a Guest Editor for many journals, including *IEEE Wireless Communications*, *International Journal of Machine Learning and Cybernetics, Information Processing* and *Management* (Elsevier), *Journal of Network and Systems Management* (Springer), *Cluster Computing* (Springer), *Internet Technology Letters*, *Transactions on Telecommunications Technologies*, *Security and Privacy*, and IEEE ACCESS. He is also an Associate Editor for *Cluster Computing*, *Security and Privacy*, IEEE ACCESS, *IET Quantum Computing*. He is a Professional Engineer Ontario (P.Eng.).